

```
In [4]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the datasets
df_worldcups = pd.read_csv('WorldCups.csv')
df_matches = pd.read_csv('WorldCupMatches.csv')
df_players = pd.read_csv('WorldCupPlayers.csv')

# Preview the datasets to understand the structure
print("WorldCups Dataset Columns:", df_worldcups.columns)
print("WorldCupMatches Dataset Columns:", df_matches.columns)
print("WorldCupPlayers Dataset Columns:", df_players.columns)

df_worldcups.head()
```

```
WorldCups Dataset Columns: Index(['Year', 'Country', 'Winner', 'Runners-Up', 'Third', 'Fourth',
    'GoalsScored', 'QualifiedTeams', 'MatchesPlayed', 'Attendance'],
    dtype='object')
WorldCupMatches Dataset Columns: Index(['Year', 'Datetime', 'Stage', 'Stadium', 'City', 'Home Team Name',
    'Home Team Goals', 'Away Team Goals', 'Away Team Name',
    'Win conditions', 'Attendance', 'Half-time Home Goals',
    'Half-time Away Goals', 'Referee', 'Assistant 1', 'Assistant 2',
    'RoundID', 'MatchID', 'Home Team Initials', 'Away Team Initials'],
    dtype='object')
WorldCupPlayers Dataset Columns: Index(['RoundID', 'MatchID', 'Team Initials', 'Coach Name', 'Line-up',
    'Shirt Number', 'Player Name', 'Position', 'Event'],
    dtype='object')
```

```
Out[4]:
```

	Year	Country	Winner	Runners-Up	Third	Fourth	GoalsScored	Qualific
0	1930	Uruguay	Uruguay	Argentina	USA	Yugoslavia	70	
1	1934	Italy	Italy	Czechoslovakia	Germany	Austria	70	
2	1938	France	Italy	Hungary	Brazil	Sweden	84	
3	1950	Brazil	Uruguay	Brazil	Sweden	Spain	88	
4	1954	Switzerland	Germany FR	Hungary	Austria	Uruguay	140	

```
In [5]: # Count the number of wins by each country
winners_count = df_worldcups['Winner'].value_counts().reset_index()
winners_count.columns = ['Country', 'Wins']

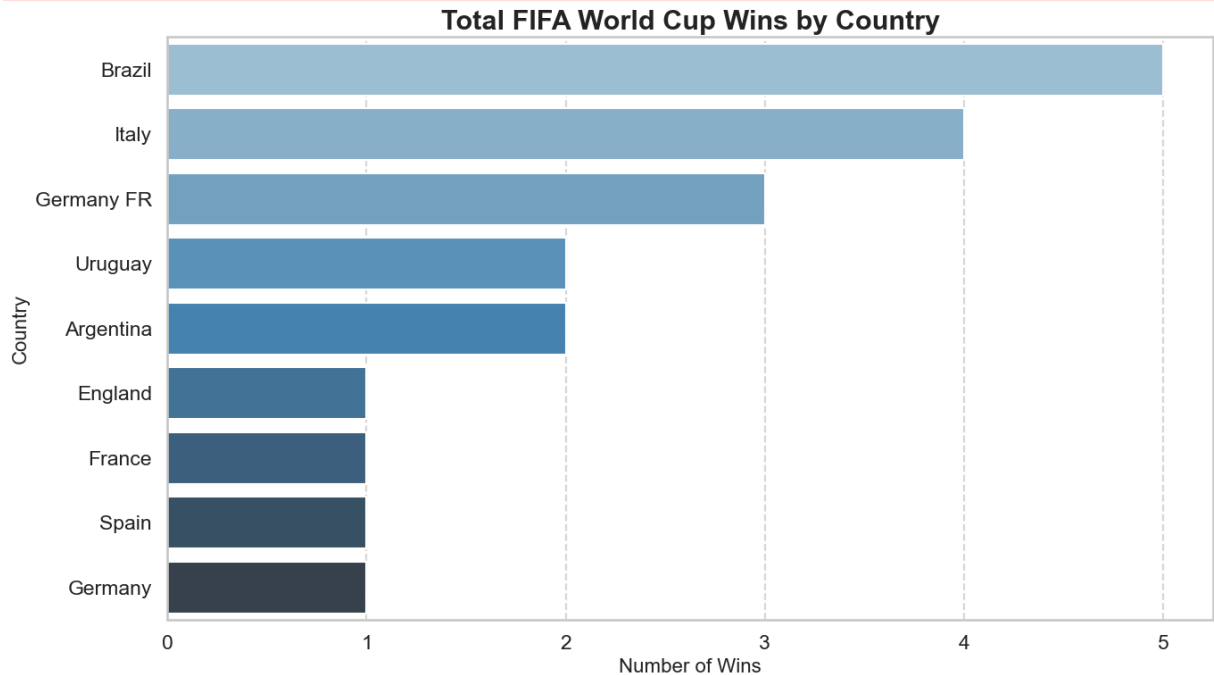
# Plot the number of wins by country
plt.figure(figsize=(14, 8))
sns.barplot(data=winners_count, x='Wins', y='Country', palette='Blues_d')
plt.title('Total FIFA World Cup Wins by Country', fontsize=22, fontweight='bold')
plt.xlabel('Number of Wins', fontsize=16)
```

```
plt.ylabel('Country', fontsize=16)
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

C:\Users\acer\AppData\Local\Temp\ipykernel_24496\3559418714.py:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=winners_count, x='Wins', y='Country', palette='Blues_d')
```



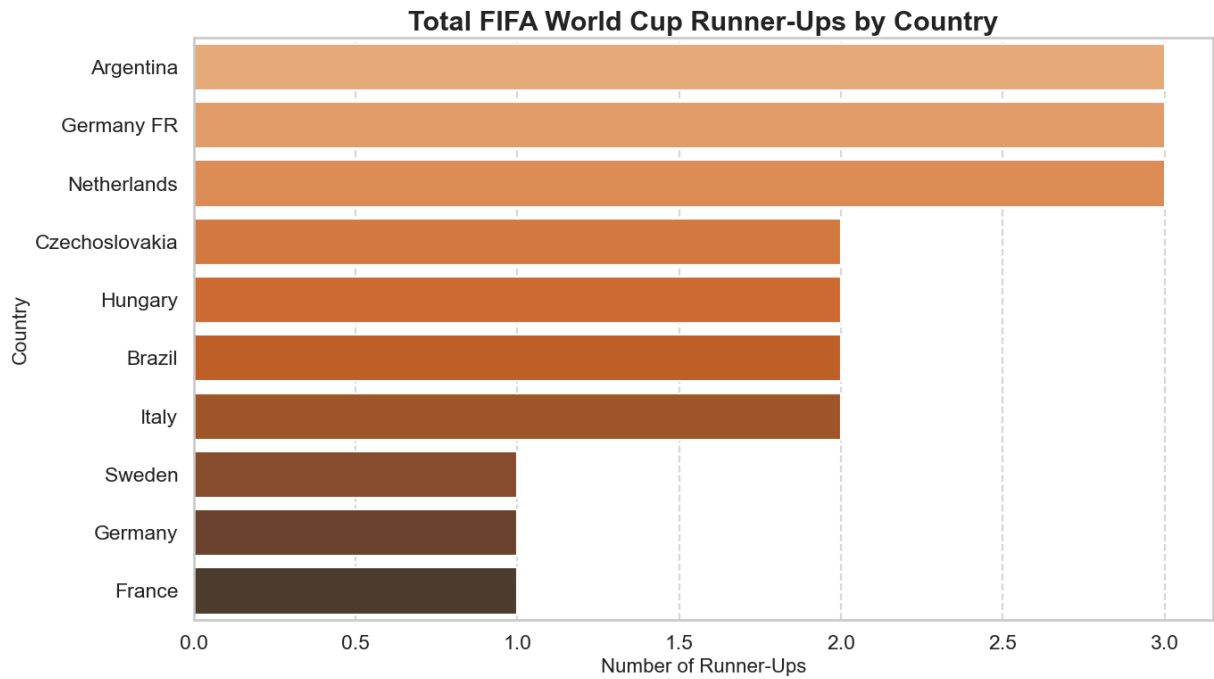
```
In [6]: # Count the number of runner-ups by each country
runners_up_count = df_worldcups['Runners-Up'].value_counts().reset_index()
runners_up_count.columns = ['Country', 'Runner-Ups']

# Plot the number of runner-ups by country
plt.figure(figsize=(14, 8))
sns.barplot(data=runners_up_count, x='Runner-Ups', y='Country', palette='Oranges_d')
plt.title('Total FIFA World Cup Runner-Ups by Country', fontsize=22, fontweight='b')
plt.xlabel('Number of Runner-Ups', fontsize=16)
plt.ylabel('Country', fontsize=16)
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

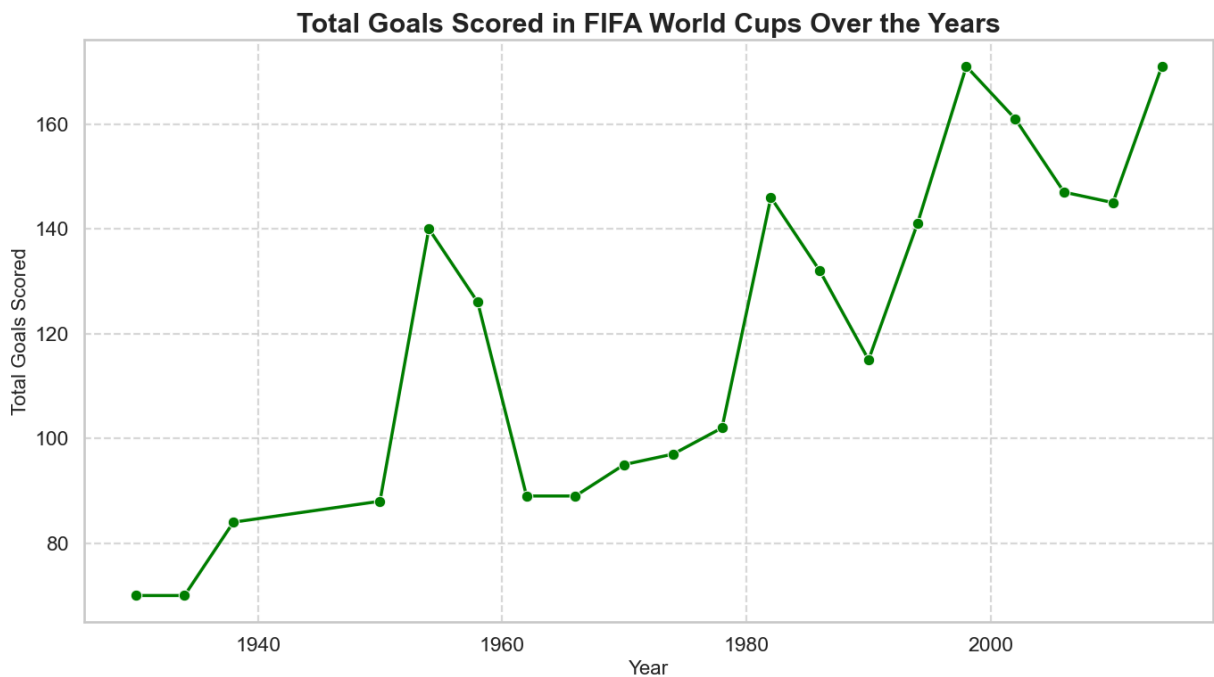
C:\Users\acer\AppData\Local\Temp\ipykernel_24496\3372030935.py:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=runners_up_count, x='Runner-Ups', y='Country', palette='Oranges_d')
```

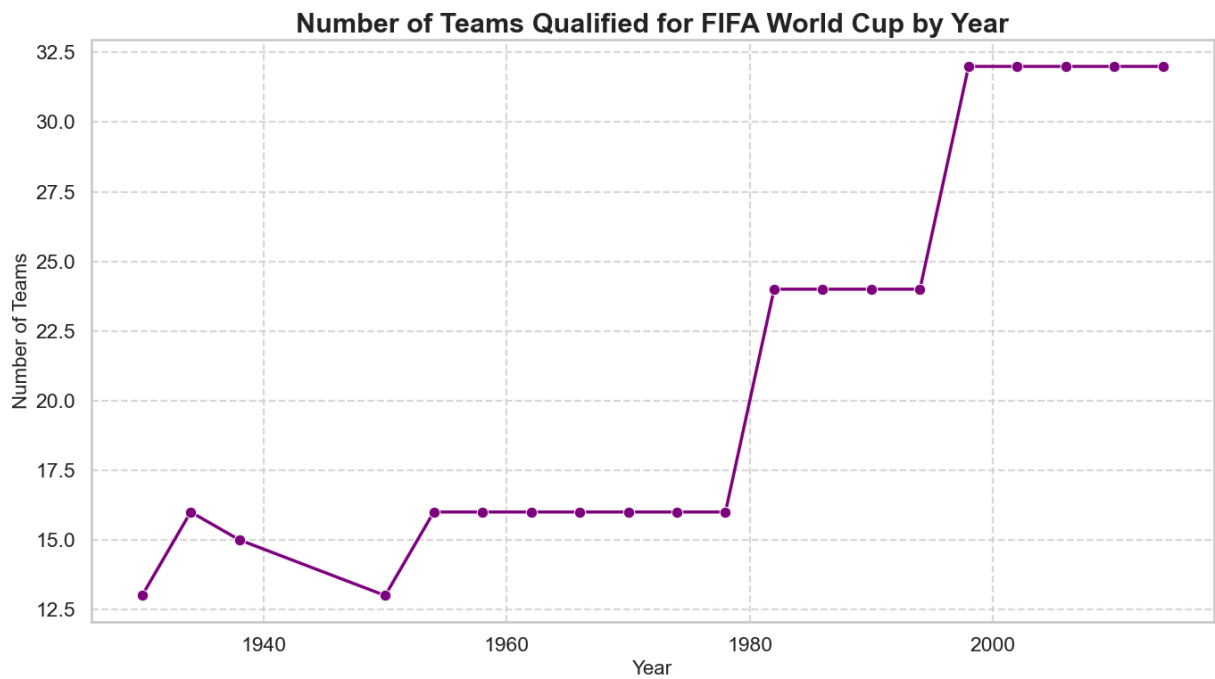


```
In [7]: plt.figure(figsize=(14, 8))
sns.lineplot(data=df_worldcups, x='Year', y='GoalsScored', marker='o', color='green')
plt.title('Total Goals Scored in FIFA World Cups Over the Years', fontsize=22, fontcolor='red')
plt.xlabel('Year', fontsize=16)
plt.ylabel('Total Goals Scored', fontsize=16)
plt.grid(linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

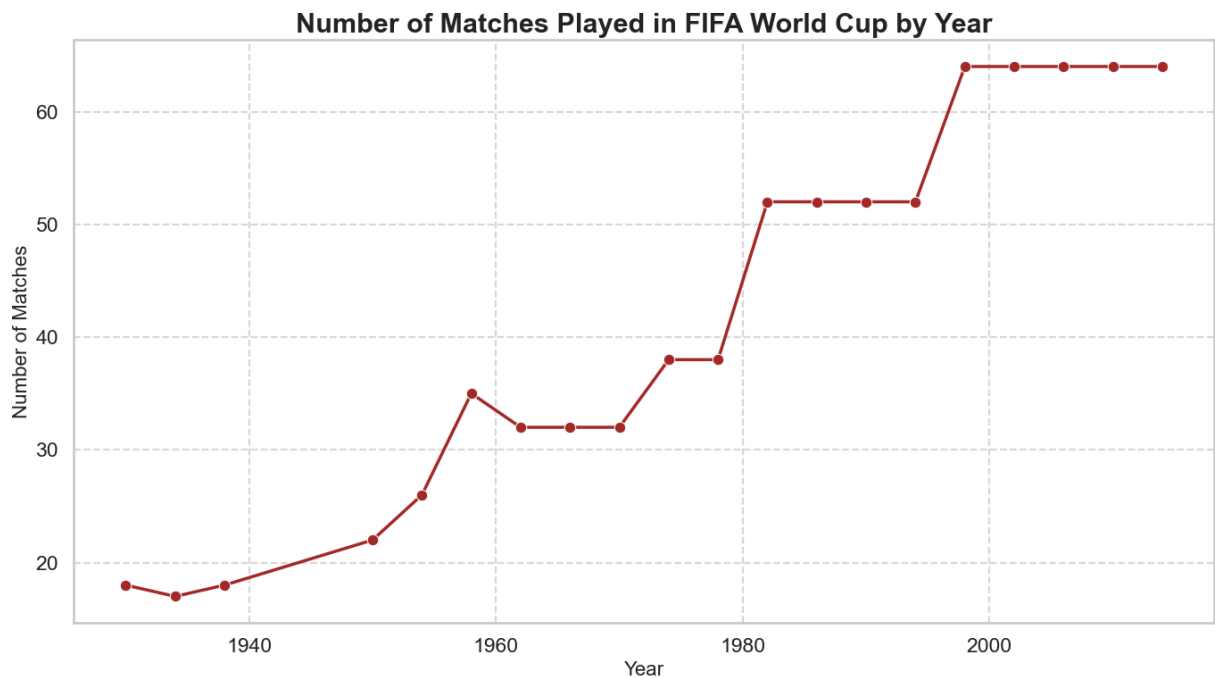


```
In [9]: plt.figure(figsize=(14, 8))
sns.lineplot(data=df_worldcups, x='Year', y='QualifiedTeams', marker='o', color='purple')
plt.title('Number of Teams Qualified for FIFA World Cup by Year', fontsize=22, fontcolor='red')
plt.xlabel('Year', fontsize=16)
plt.ylabel('Number of Teams', fontsize=16)
```

```
plt.grid(linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



```
In [10]: plt.figure(figsize=(14, 8))
sns.lineplot(data=df_worldcups, x='Year', y='MatchesPlayed', marker='o', color='brown')
plt.title('Number of Matches Played in FIFA World Cup by Year', fontsize=22, fontw
plt.xlabel('Year', fontsize=16)
plt.ylabel('Number of Matches', fontsize=16)
plt.grid(linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

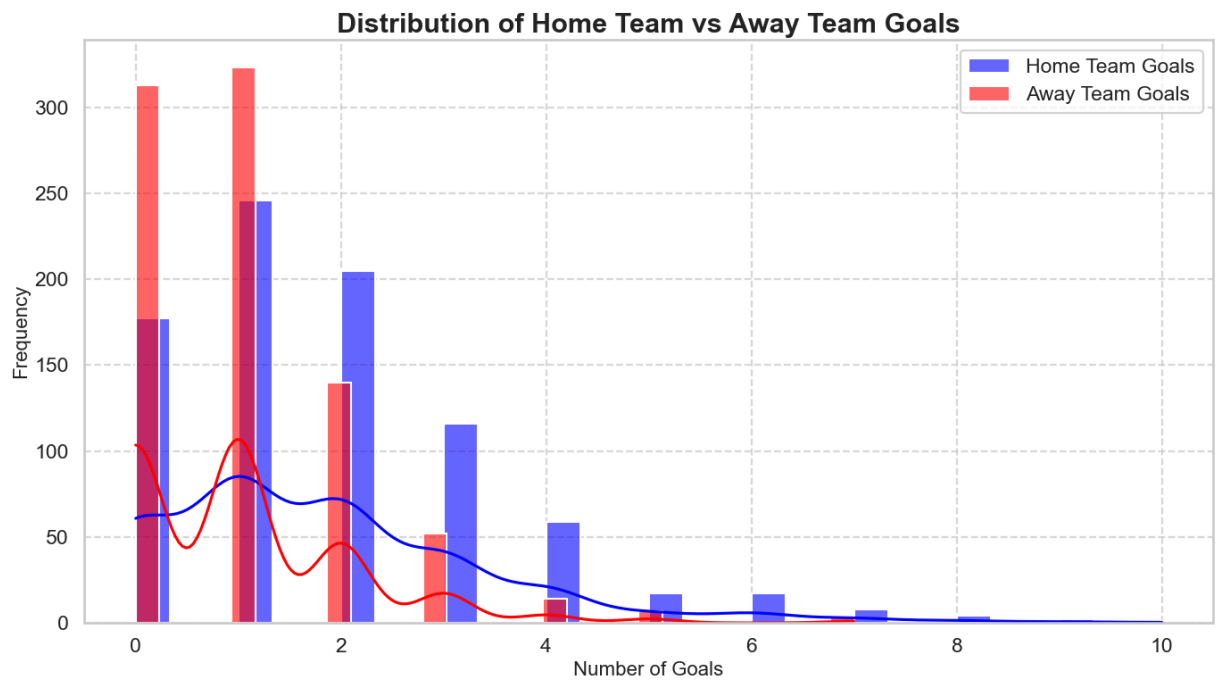


```
In [11]: plt.figure(figsize=(14, 8))

# Plot home team goals
sns.histplot(df_matches['Home Team Goals'], kde=True, color='blue', label='Home Te

# Plot away team goals
sns.histplot(df_matches['Away Team Goals'], kde=True, color='red', label='Away Tea

plt.title('Distribution of Home Team vs Away Team Goals', fontsize=22, fontweight=
plt.xlabel('Number of Goals', fontsize=16)
plt.ylabel('Frequency', fontsize=16)
plt.legend()
plt.grid(linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



In []: